How do you write a case statement in Pandas? Can you provide an example?

Authored by

stats writer

June 30, 2024

A case statement in Pandas is a conditional statement used to create a new column in a dataset based on certain conditions. It is used to transform data and make it more meaningful for analysis. To write a case statement in Pandas, you can use the "np.where()" function, which takes in three arguments - a condition, a value to be assigned if the condition is true, and a value to be assigned if the condition is false. An example of a case statement in Pandas is as follows:

df = np.where(df >= 18, 'adult', 'minor')

This statement creates a new column called "new_column" in the dataframe "df" and assigns the value "adult" if the age of a person is greater than or equal to 18, and "minor" if the age is less than 18. This allows for easy categorization and analysis of data based on age.

Write a Case Statement in Pandas (With Example)

A case statement is a type of statement that goes through conditions and returns a value when the first condition is met.

The easiest way to implement a case statement in a Pandas DataFrame is by using the NumPy where() function, which uses the following basic syntax:

df = np.where(df<9, 'value1',
np.where(df<12, 'value2',
np.where(df<15, 'value3', 'value4')))

This particular function looks at the value in the column called col2 and returns:

"value1" if the value in col2 is less than 9"value2" if the

value in col2 is less than 12"value3" if the value in col2 is less than 15"value4" if none of the previous conditions are true

The following example shows how to use this function in practice.

Example: Case Statement in Pandas

Suppose we have the following pandas DataFrame:

```
import pandas as pd
import numpy as np

#create DataFrame
df = pd.DataFrame({'player': ,
'points': })

#view DataFrame
df
```

player points
0 1 6
1 2 8
2 3 9
3 4 9

4 5 12

5 6 14

6 7 15

7 8 17

8 9 19

9 10 22

We can use the following syntax to write a case statement that creates a new column called class whose values are determined by the values in the points column:

```
#add 'class' column using case-statement logic
df = np.where(df<9, 'Bad',
np.where(df<12, 'OK',
np.where(df<15, 'Good', 'Great')))

#view updated DataFrame
df
```

player points class

0 1 6 Bad

1 2 8 Bad

2 3 9 OK

3 4 9 OK

4 5 12 Good

5 6 14 Good

6 7 15 Great

7 8 17 Great

8 9 19 Great

9 10 22 Great

The case statement looked at the value in the points column and returned:

"Bad" if the value in the points column was less than 9"OK" if the value in the points column was less than 12"Good" if the value in the points column was less than 15"Great" if none of the previous conditions are true

Note: You can find the complete documentation for the NumPy where() function .

Additional Resources

The following tutorials explain how to perform other common tasks in Pandas: